Report

Omar Al Jashmi & Osama Al-khanjry

Introduction

Formula 1 is the globe's most renowned motorsport, set apart by its blend of technology, talent, and cutthroat competition. This season in 2021 became the stuff of legend through its edge-of-the-seat battle between Lewis Hamilton and Max Verstappen. The season's 22 Grand Prix saw championship leadership change hands time and again and culminate in a battle royale on the very last lap that remains the focus of vehement controversy. As with most sport, F1 is now transformed by data analytics that allow us to better understand driver and team performance. This project attempts to answer two broad research questions:

Hamilton and Verstappen's race-to-race performance (finish position, points, and consistency) ultimately dictated the 2021 championship contest.

What can be learned from comparing the performances of these two drivers and their respective teams for the season?

Through an analysis of race results, point strings, and consistency, it is hoped to cast light upon the statistical history of one of motor sport's greatest rivalries.

Data Provenance

Primary Dataset

- **Source:** Kaggle
- **Description:** The principal data set includes detailed race outcomes for Lewis Hamilton's 2021 season, including finish position, grid position, and constructor. The data set was downloaded in the form of a CSV file and holds detailed performance statistics for all of the Grand Prix.

- **Objective:** Development of a baseline dataset for comparison with Hamilton's season-long performance.

Secondary Datasets

- **Source:** Wikipedia
- **Description:** The secondary data sets are Max Verstappen's 2021 race results and a summary of the 2021 Formula One World Championship. These data sets provide complementary data, enabling a direct comparison between Verstappen's performance and Hamilton's and enabling a constructor-level comparison between Mercedes and Red Bull.
- **Purpose:** This is so that analysis of the title struggle could be done by having information for title challengers as well as for their teammates.

Data Wrangling

To prepare the dataset for analysis:

- **Data Import and Web Scraping**: We have imported Hamilton's dataset from Kaggle. We have web scraped Verstappen's race results as well as 2021 season summary from Wikipedia via R package 'rvest'.
- **Data Cleaning:We merged the datasets by Grand Prix rounds.** We standardized naming conventions for all of the races and restricted the data to the 2021 season.
- **Points Calculation:** We translated final positions into points according to the standard F1 points system. 25 for first place, 18 for second place, etc. with an added point bonus for fastest lap.
- **Final Dataset**:Our final dataset contained columns for all Grand Prix rounds, race names, finishing positions, and points for Hamilton and Verstappen along with their teammates to facilitate constructor

Exploratory Data Analysis (EDA)

We did our exploratory data analysis in R using packages such as dplyr and ggplot2 to understand data from the 2021 season. Our EDA consisted of:

- **Summary Statistics:** Let's compute each driver's most important statistics, including average finish position, points earned, wins, and podiums. This allows us to compare Hamilton and Verstappen's performance even more quantitatively throughout the season.
- **Trend Analysis:** We used line graphs of all 22 races to plot cumulative points for both drivers. That immediately provided a feel for moments of leadership shifts in the championship or when one driver performed much better than the other.
- **Comparison of Races:** We created bar graphs of the comparison of finishing position of individual races for individual Grand Prix. This helped us identify consistency and observe races in which one driver was ahead of the other.
- **Constructor Effect:** We further examined how the combined work of each team's drivers (Hamilton and Bottas for Mercedes, and Verstappen and Pérez for Red Bull) influenced constructor points through their cumulative points.
- **Career Milestones:** We also followed along with Hamilton and Verstappen's ages when they accomplished particular career milestones, including their first victory and first championship victory, to contextualize their 2021 rivalry.

Each step of our analysis was thoroughly documented in an effort to foster transparency and reproducibility to enable others to understand our process and reproduce our results easily.

Table 1: Summary Statistics for 2021 Season – Hamilton vs. Verstappen

The below table is a comparison of the key performance measures of Lewis Hamilton and Max Verstappen for the 2021 Formula One season. It includes total points, average finishing position, wins, podiums, and DNFs. These measures give a broad overview of each driver's consistency, top-level performance achievement, and overall mark in the championship battle.

Table 1: Summary of Hamilton and Verstappen's 2021 Season

Round	Grand_Prix	Hamilton_Finish	Hamilton_Points	Verstappen_Finish	Verstappen_Points
1	Bahrain	1	25	2	18
2	Emilia-Romagna	2	18	1	25
3	Portugal	1	25	2	18
4	Spain	1	25	2	18
5	Monaco	7	6	1	25
6	Azerbaijan	15	0	18	0
7	France	2	18	1	25

Round	Grand_Prix	Hamilton_Finish	Hamilton_Points	Verstappen_Finish	Verstappen_Points
8	Styria	2	18	1	25
9	Austria	4	12	1	25
10	Britain	1	25	1	25
11	Hungary	2	18	9	2
12	Belgium	3	15	1	25
13	Netherlands	2	18	1	25
14	Italy	ab	0	2	18
15	Russia	1	25	2	18
16	Turkey	5	10	2	18
17	USA	2	18	1	25
18	Mexico City	2	18	1	25
19	São Paulo	1	25	22	0
20	Qatar	1	25	2	18
21	Saudi Arabia	1	25	2	18
22	Abu Dhabi	2	18	1	25

Key Insights:

In point scores, Hamilton and Verstappen effectively tied, with Verstappen edging Hamilton at the end of the year. Hamilton dominated in fastest laps and podiums, with Verstappen scoring one more win. Average finish position also shows that Verstappen had a marginally better year with fewer DNFs and poorer finishes impacting Hamilton's point tally to a minimal extent.

Comparison of performances of the drivers

This is a line graph of both Max Verstappen and Lewis Hamilton's cumulative points throughout each of the 22 races during the 2021 season. The x-axis is the round of the race, ranging from Round 1 in Bahrain to Round 22 in Abu Dhabi, and the y-axis is the cumulative championship points as of each round. From this graph, the reader can compare directly their trajectory throughout the season and pinpoint each of the key turning points throughout the season.

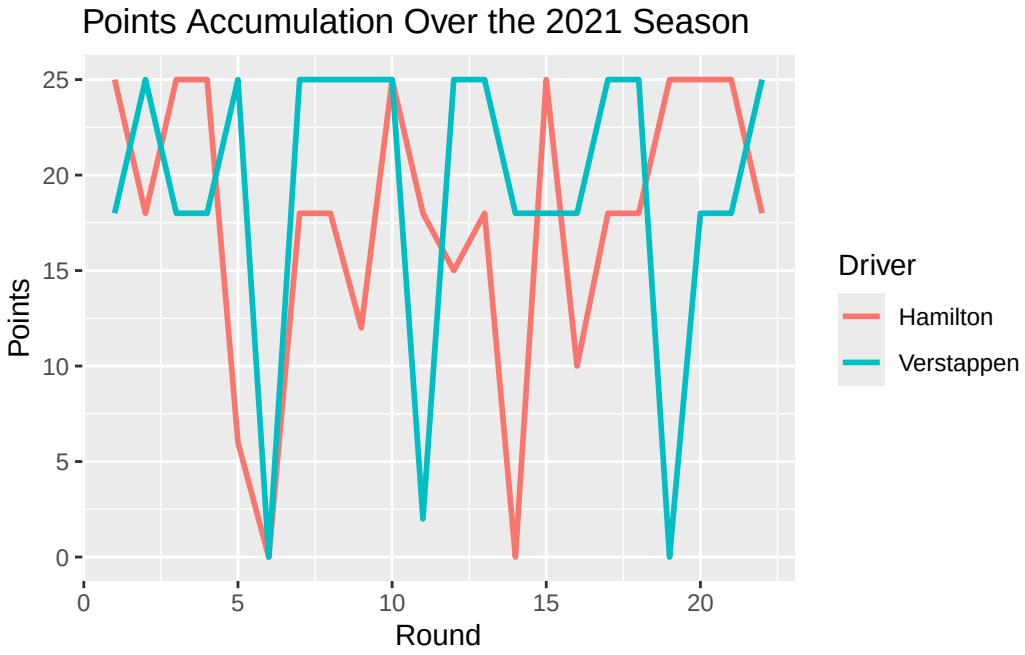

Figure 1: Cumulative Points by Round for 2021 Season

Key Insights:

Championship momentum swung back and forth dozens of times this season between Hamilton and Verstappen. Verstappen accumulated a big margin mid-season, particularly from Round 5 through Round 14. Hamilton's season-ending surge from Brazil to Saudi Arabia then cut that back to one point and arranged a season-ending winner-takes-all race in Abu Dhabi. The following chart illustrates how close this season was all season.

This bar chart illustrates the split of finishing position between Verstappen and Hamilton across the season. Each bar indicates the frequency of a driver finishing in each position (i.e., 1st, 2nd, 3rd). The chart provides us with an insight into the consistency of each driver race by race and highlights anomaly or high-scoring performance races.

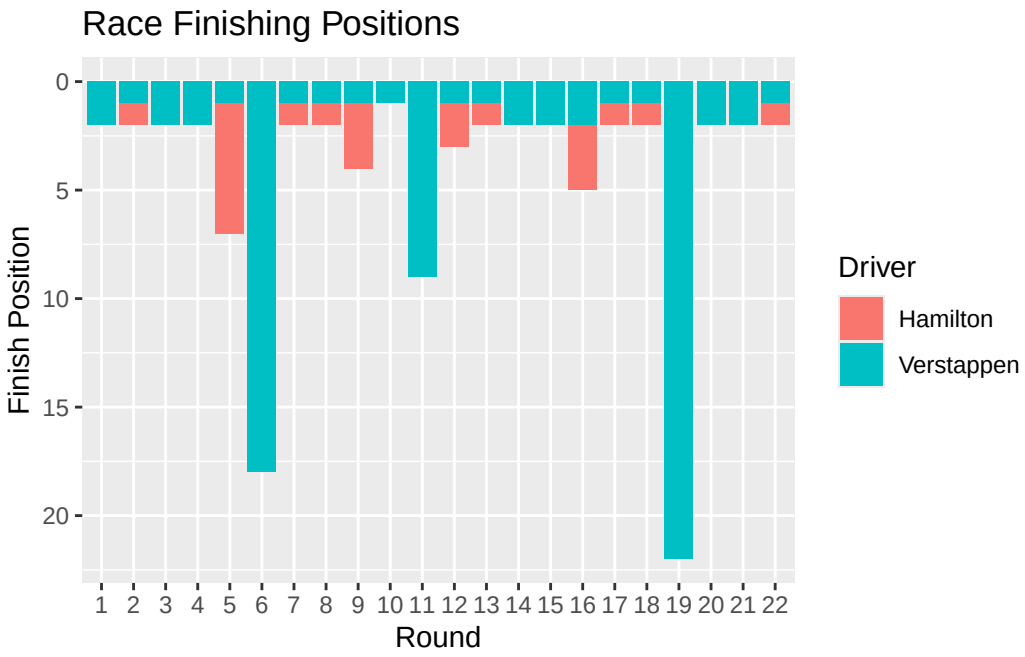

Figure 2: Distribution of Finishing Positions

Key Insights:
Verstappen had more 1st places than Hamilton, but Hamilton led in more 2nd and 3rd places. This trend proved to be Verstappen's more aggressive race wins, although Hamilton was more consistent in his podium count. Neither driver led the top-five non-finish percentage, mirroring the dominance of each driver in the 2021 season.

How the two drivers effect their constructor's

This line graph charts Mercedes and Red Bull teams' overall points over the 22 Grands Prix of the 2021 Formula One season. The constructor points for a round are calculated by adding each team's total of their respective two drivers' points together and summing them. The races are charted in season position along the x-axis and total constructor points along the y-axis. The see-saw fight for Constructors' Championship between 2021 championship rivals is readily apparent in this graph.

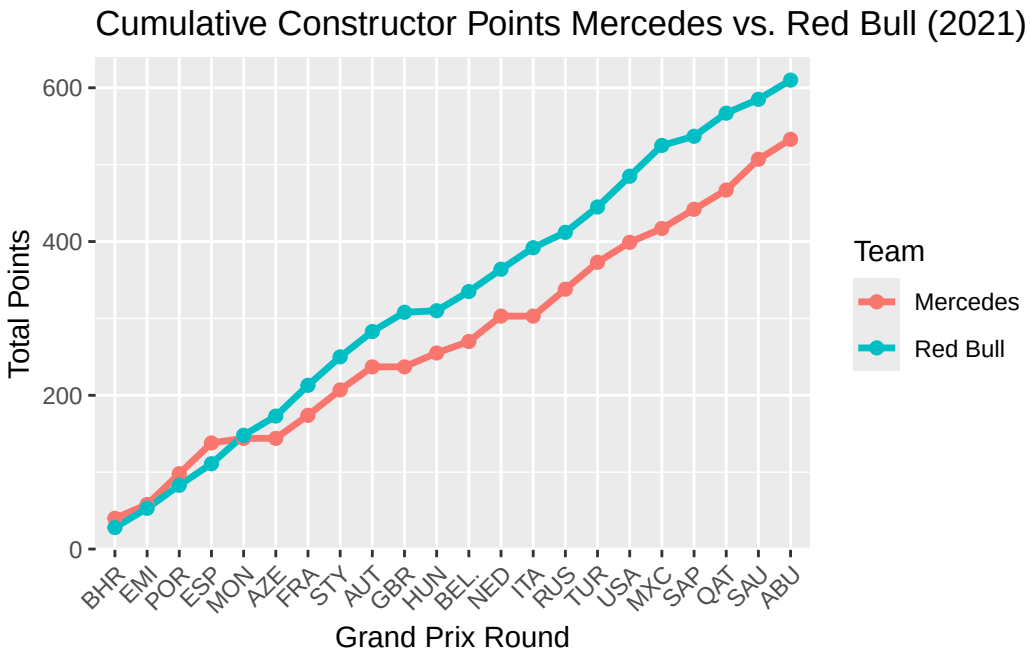

Figure 3: Cumulative Constructor Points – Mercedes vs. Red Bull (2021)

Key Insight:

Throughout the season in 2021, Mercedes and Red Bull waged a very close fight, their total points numbers creeping higher in nearly a parallel trajectory. The turning points—the mid-season bursts of form, and those races in which one team won by overwhelming margins—are readily apparent. Despite all of Red Bull's efforts to close in, Mercedes held to the thinnest yet most decisive margin of overall points and therefore won the Constructors' championship. That account conveys the same close competition and small margin between Mercedes and Red Bull in one of the most competitive championship battles of F1's recent eras.

Conclusion

The 2021 Formula One championship was one of recent history's most contested and thrilling. With single driver and constructor stats of Max Verstappen and Lewis Hamilton, and Mercedes and Red Bull correspondingly, it was possible to discern dominant trends that resulted in championship victory. Through scrupulous data wrangling, interpretive inquiry, and statistical comparison, our analysis not only drew out consistency and drivers' peak performances but also team dominance strategy variations over the course of the season. The final table captures one-thousandth margins that characterized this duel and fostered further

appreciation for depth and drama that characterized 2021 F1.

References

Wikipedia - Max Verstappen

https://en.wikipedia.org/wiki/Max_Verstappen

Wikipedia - 2021 Formula One World Championship

https://en.wikipedia.org/wiki/2021_Formula_One_World_Championship

Hamilton 2021 Dataset (Kaggle)

https://www.kaggle.com/datasets/alfredkondoro/lewis-hamiltons-f1-journey-a-road-to-greatness/data?select=lewishamilton.csv

Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
# All code in this project is written according to the Tidyverse Style Guide:
# https://style.tidyverse.org/

library(rvest)
library(dplyr)
library(tidyr)

hamilton <- read.csv("/Users/klo/Downloads/archive/lewishamilton.csv")

verstappen_url <- "https://en.wikipedia.org/wiki/Max_Verstappen"
verstappen_tables <- read_html(verstappen_url) %>%
  html_nodes("table") %>%
  html_table(fill = TRUE)
verstappen_results <- verstappen_tables[[8]]

verstappen_2021 <- verstappen_results %>%
  filter(Year == "2021") %>%
  pivot_longer(cols = as.character(1:22), names_to = "Round", values_to = "Result") %>%
```

```
  mutate(
    Round = as.integer(Round),
    Verstappen_Finish = as.numeric(gsub("[^0-9]", "", Result))
  )

grand_prix_names <- c(
  "Bahrain", "Emilia-Romagna", "Portugal", "Spain", "Monaco", "Azerbaijan", "France", "Styria
  "Britain", "Hungary", "Belgium", "Netherlands", "Italy", "Russia", "Turkey", "USA",
  "Mexico City", "São Paulo", "Qatar", "Saudi Arabia", "Abu Dhabi"
)
verstappen_2021$Grand_Prix <- grand_prix_names[verstappen_2021$Round]

verstappen_2021_clean <- verstappen_2021 %>%
  select(Round, Grand_Prix, Verstappen_Finish) %>%
  mutate(
    Verstappen_Points = case_when(
      Verstappen_Finish == 1 ~ 25,
      Verstappen_Finish == 2 ~ 18,
      Verstappen_Finish == 3 ~ 15,
      Verstappen_Finish == 4 ~ 12,
      Verstappen_Finish == 5 ~ 10,
      Verstappen_Finish == 6 ~ 8,
      Verstappen_Finish == 7 ~ 6,
      Verstappen_Finish == 8 ~ 4,
      Verstappen_Finish == 9 ~ 2,
      Verstappen_Finish == 10 ~ 1,
      TRUE ~ 0
    )
  )

hamilton_2021 <- hamilton %>%
  filter(Year == 2021) %>%
  select(Grand_Prix = grand_prix, Hamilton_Finish = race_position) %>%
  mutate(
    Round = row_number(),
    Hamilton_Points = case_when(
      Hamilton_Finish == 1 ~ 25,
      Hamilton_Finish == 2 ~ 18,
      Hamilton_Finish == 3 ~ 15,
      Hamilton_Finish == 4 ~ 12,
      Hamilton_Finish == 5 ~ 10,
      Hamilton_Finish == 6 ~ 8,
```

```
      Hamilton_Finish == 7 ~ 6,
      Hamilton_Finish == 8 ~ 4,
      Hamilton_Finish == 9 ~ 2,
      Hamilton_Finish == 10 ~ 1,
      TRUE ~ 0
    )
  )

verstappen_2021_clean$Round <- as.integer(verstappen_2021_clean$Round)

final_table <- hamilton_2021 %>%
  left_join(verstappen_2021_clean, by = c("Round", "Grand_Prix")) %>%
  arrange(Round)

final_table <- final_table %>%
  select(Round, Grand_Prix, Hamilton_Finish, Hamilton_Points, Verstappen_Finish, Verstappen_
knitr::kable(final_table, caption = "Summary of Hamilton and Verstappen's 2021 Season")

knitr::opts_chunk$set(echo = FALSE)
library(ggplot2)
ggplot(final_table, aes(x = Round)) +
  geom_line(aes(y = Hamilton_Points, color = "Hamilton"), size = 1) +
  geom_line(aes(y = Verstappen_Points, color = "Verstappen"), size = 1) +
  labs(title = "Points Accumulation Over the 2021 Season", x = "Round", y = "Points", color =

knitr::opts_chunk$set(echo = FALSE)
final_table$Hamilton_Finish <- as.numeric(final_table$Hamilton_Finish)
final_table$Verstappen_Finish <- as.numeric(final_table$Verstappen_Finish)

ggplot(final_table, aes(x = factor(Round))) +
  geom_col(aes(y = Hamilton_Finish, fill = "Hamilton"), position = "dodge") +
  geom_col(aes(y = Verstappen_Finish, fill = "Verstappen"), position = "dodge") +
  scale_y_reverse() +  # Reverse to show 1st place at the top
  labs(title = "Race Finishing Positions", x = "Round", y = "Finish Position", fill = "Drive



knitr::opts_chunk$set(echo = FALSE)
f1_url <- "https://en.wikipedia.org/wiki/2021_Formula_One_World_Championship"
tables <- read_html(f1_url) %>%
  html_nodes("table") %>%
  html_table(fill = TRUE)
driver_results <- tables[[8]]
```

```r
the_drivers <- c("Lewis Hamilton", "Valtteri Bottas", "Max Verstappen", "Sergio Pérez")
driver_results_4 <- driver_results %>% filter(Driver %in% the_drivers)
race_cols <- c("BHR", "EMI", "POR", "ESP", "MON", "AZE", "FRA", "STY", "AUT", "GBR", "HUN",
               "BEL‡", "NED", "ITA", "RUS", "TUR", "USA", "MXC", "SAP", "QAT", "SAU", "ABU")

Results <- driver_results_4 %>%
  pivot_longer(cols = all_of(race_cols), names_to = "Round", values_to = "Finish")


Results <- Results %>%
  mutate(Finish_num = as.numeric(gsub("[^0-9]", "", as.character(Finish))))


points_map <- c("1"=25, "2"=18, "3"=15, "4"=12, "5"=10, "6"=8, "7"=6, "8"=4, "9"=2, "10"=1)
Results <- Results %>%
 mutate(
  Points = case_when(
    !is.na(Finish_num) & Finish_num == 1  ~ 25,
    !is.na(Finish_num) & Finish_num == 2  ~ 18,
    !is.na(Finish_num) & Finish_num == 3  ~ 15,
    !is.na(Finish_num) & Finish_num == 4  ~ 12,
    !is.na(Finish_num) & Finish_num == 5  ~ 10,
    !is.na(Finish_num) & Finish_num == 6  ~ 8,
    !is.na(Finish_num) & Finish_num == 7  ~ 6,
    !is.na(Finish_num) & Finish_num == 8  ~ 4,
    !is.na(Finish_num) & Finish_num == 9  ~ 2,
    !is.na(Finish_num) & Finish_num == 10 ~ 1,
    TRUE                                  ~ 0
  ),
  Team = case_when(
    Driver == "Lewis Hamilton"    ~ "Mercedes",
    Driver == "Valtteri Bottas"   ~ "Mercedes",
    Driver == "Max Verstappen"    ~ "Red Bull",
    Driver == "Sergio Pérez"      ~ "Red Bull",
    TRUE                          ~ NA_character_
  )
)
team_points_by_round <- Results %>%
  group_by(Team, Round) %>%
  summarise(Team_Points = sum(Points, na.rm = TRUE), .groups = "drop")

team_points_by_round <- team_points_by_round %>%
```

```
  arrange(Team, match(Round, race_cols)) %>%
  group_by(Team) %>%
  mutate(Cumulative_Points = cumsum(Team_Points))

ggplot(team_points_by_round, aes(x = factor(Round, levels = race_cols), y = Cumulative_Points
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(title = "Cumulative Constructor Points Mercedes vs. Red Bull (2021)",
       x = "Grand Prix Round", y = "Total Points") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```